Defence Research and Development Canada
Recherche et développement pour la défense Canada

Microarray Genomic Systems Development

V. Lam, M. Crichton, T. Dickinson Laing, and D.C. Mah
Canada West Biosciences Inc.

Contract Scientific Authority: B. Ford, DRDC Suffield

Defence R&D Canada

Contract Report

DRDC Suffield CR 2009-145

June 2008

Canada

Microarray Genomic Systems Development

V. Lam, M. Crichton, T. Dickinson Laing, and D.C. Mah
Canada West Biosciences Inc.

Canada West Biosciences Inc.
5429 60th Street
Camrose, AB
T4V 4G9

Contract Number: W7702-04-R053/001/EDM

Contract Scientific Authority: B. Ford (403-544-4612)

Defence R&D Canada – Suffield

Contract Report
DRDC Suffield CR 2009-145
June 2008

Principal Author

Victor Lam

Research Team Member

Approved by

Approved for release by

Abstract

In order to identify and characterize microbes using the currently available methods, information on the organism's genetic material is required. Recombinant microbes which contain novel genetic material would not have such information and therefore avoid detection. Previous work has been done to develop a system with array-based genomic fingerprinting technology, which should help to identify the new strains, detect the presence of novel genetic material, and measure gene expressions. The next phase is the development of a test system for rapid species identification in addition to the oligonucleotide fingerprint. The test organisms for this work were *Bacillus* bacteria (11 species), *Escherichia coli TOP10* (7 strains), and *Geobacillus stearothermophilus*. Using standard molecular biology methods, we isolated genomic DNA, digested the DNA to reduce complexity, labelled it with fluorescent dyes, and hybridized the labelled DNA to two types of microarrays, Human Operon 21K chips containing 23,232 features and Bacterial genomic chips containing 5,280 features. The hybridization data was then analyzed with ChromaBlast, an useful analytic tool in Excel, which normalized columnar data, sorted the data into user-selectable range-driven bins, developed colour heat maps from the data, and then output the heat map and bin assortment for review. When the data patterns on the colour heat maps were filtered and sorted, bacteria in different genera could be discriminated with high confidence in certain subsets of the features. However, species or strain differentiations of the test organisms were not as evident in this work. A multipathogen chip was designed to further investigate species and strain differentiation. Due to time limitation on the contract, only one or two sample chips from each test organism has been included thus far. In the next phase, data from more replicate microarray chips should be included in the set of hybridization data.

This page intentionally left blank.

Executive summary

Microarray Genomic Systems Development

Lam, V.; Crichton, M.; Dickinson Laing, T.; Mah, D.C.; DRDC Suffield CR 2009-145; Defence R&D Canada – Suffield; June 2008.

Introduction: Conventional methods of microbe identification require genomic information of the suspect organism, utilizing molecular biology techniques microbes could be identified to the species level. However, since recombinant organisms containing novel genetic content, the existing library of genomic sequences might not have information on these new organisms. Therefore recombinant organisms are harder to identify and characterize, and they are more likely to evade detection by existing techniques. The new method being developed here is to use a large panel of non-specific oligonucleotides in a microarray hybridization format, to interrogate genomic DNA from a series of test organisms. This assay would output reproducible patterns of hybridization on the microarray slide. These patterns, as "fingerprints", would be stored to compile a library of identified bacterial strains for computational comparisons.

Results: Different species of bacteria, including *Escherichia coli*, *Bacillus* bacteria, and *Geobacillus stearothermophilus* produce qualitatively different hybridization patterns on the microarrays. Analysis included comparison of hybridization patterns on the colour heat map output from the hybridization data. The colour heat map was set up to show greater gene expressions at the feature spots in "hotter" colours, i.e. near the red end on the spectrum, while lower gene expressions at feature spots were in "cooler" colours, i.e. near the blue end of the spectrum. Dark spots on the microarray were represented by gray colour bins on the colour heat map. When the colour heat maps were filtered, at specific features the colours of *Bacillus* bacteria were "hotter" (red) than the *Escherichia coli* strains (blue), which suggested gene expressions at these features were greater for the *Bacillus* bacteria than the *Escherichia coli* strains. The hybridization patterns between all the *Bacillus* species, and between the *Escherichia coli* strains were visibly different as well, but features where there were differentiating patterns were not as evident. In this work species differences were detected with lower confidence by inspection of array patterns.

Significance: It has been demonstrated that differential hybridization patterns were indeed produced by different bacterial species. Hybridization patterns of recombinant organisms with novel genetic content should be notably different from those observed in library strains. These recombinant organisms should be detectable even when there was no previous information on their genetic content. Although species and strain differentiations were less evident in this work, from what we saw based on one set of microarray chips for the test organisms, the potential for microbe species and strain differentiation do exist.

Future plans: For some test organisms there were multiple microarray chips produced, all of these replicate chips shall have their results exported and included in the hybridization data as well. More microarray chips should be produced to compile multiple replicates for each test organism, with the average values to get more accurate representations of the output data. Furthermore, existing data of other bacterial species from previous experiments could be added to the hybridization data set for the purpose of comparison in future as well.

This page intentionally left blank.

Table of contents

List of figures

List of tables

Acknowledgements

The authors would like to thank Dr. Barry Ford, Mr. Michael McWilliams, and Mr. Yimin Shei at DRDC Suffield for invaluable advice and contribution in all portions of the microarray work. The authors would also like to thank Ms. Catharine Richardson for her efforts in developing the ChromaBlast Microsoft Excel add-in tool.

The authors would like to thank the Gene Array Facility at the Prostate Cancer Research Centre in Vancouver General Hospital for providing us with the spotted Human Genome and Bacterial arrays.

Introduction

The potential threat of novel pathogenic organisms with increased virulence and other modifications being used in biological warfare and bioterrorism do exists. The current methods of microbial detection and identification have certain requirements, such as the genetic and phenotypic data related to the organism in question. Recombinant organisms with novel genetic material by natural acquisition or artificial modification, which may confer novel phenotype, theoretically would fall outside the scope of the existing identification methods. For the recombinant organisms the genetic or phenotypic data is either not available or they are misidentified as data from the existing organisms.

The current methods used in microbial identification such as RFLP, PCR, comparative genomic hybridization are able to detect the presence or absence of genetic sequences which have already been profiled and included in the reference genomes [1]. However, the current methods have certain flaws which limit their usefulness in detecting recombinant pathogens, such as the risk of them being mistakenly identified as existing organisms, the lack of multiplexing assays for recombinants, and the necessity of knowing the organisms' DNA sequences for the DNA-based tests. Most importantly, it is difficult to screen for recombinant organisms with novel properties from other microbes such as antibiotic resistance or toxins. In this work, utilizing microarrays a large amount of assays could be performed simultaneously with a single sample. It also allows the comparison of many DNA sequences to known library strains, and could provide rapid identification of the unknown microbe to any biological threat. Microarrays are also used in the studies of differentially gene expressions and host-microbe interactions [2-5].

A microbial identification method is being developed utilizing microarray chips printed with known genetic targets, such as known toxin, virulence and host strain genes, which should help to identify the unknown microbial sample under the assumption that it will be matched with known species or strain. However, recombinant organism with alteration to its genetic material could fall outside the detecting range, which is why the chips are also printed with random defined oligonucleotides which do not match to any known genetic target, thus every organism should hybridize to these oligonucleotides in special patterns, yielding unique genomic fingerprints for all organisms. Among the thousands of features on the hybridization patterns, many reference points could be found in the microarray fingerprints. Utilizing statistical or cluster analysis [6], and the hybridization patterns from unknown samples digitized, the data could be compared to a library of fingerprints from known microbial species and strains for identification purpose [7-10]. Each genomic fingerprint may contain multiple subpatterns which could be filtered according to relative intensity, thus genomic fingerprinting on microarray platforms relies heavily on bioinformatics during analysis [11-13].

In this work, genomic fingerprinting was done by hybridizing oligonucleotides to labelled genomic DNA from a set of test samples, including eleven *Bacillus* species, *Geobacillus stearothermophilus*, and seven *Escherichia coli TOP10* strains. Microbial species and strain differentiation could be demonstrated at different confidence levels by highlighting and comparing specific areas on the hybridization patterns.

Materials and Methods

Genomic DNA Extraction

The method used in genomic DNA extraction from *Bacillus* species and *Geobacillus stearothermophilus* were essentially the same as for the *E. coli* DNA.

Except where otherwise noted, reagents used were obtained from SIGMA (Oakville, ON). DNA from *Escherichia coli* (strains *TOP10*) grown overnight at 37 ºC in 50 ml of Luria broth (LB) (50 µl cell pellet in LB broth, shaken at 250 rpm) was extracted by first pelleting the bacteria at 2000 rpm for 25 minutes. For Gram-negative bacteria, treat the cell pellet with 5 times volume of water; And for Gram-positive bacteria, treat the cell pellet with 5 times volume of lysozyme solution (5 mg/ml lysozyme in 50 mM glucose, 10 mM EDTA, and 25 mM Tris HCl, pH 8.0) at 37 ºC for 30 minutes. This was followed by 20 ml of DNAzol (Invitrogen, Carlsbad, CA), mixed gently with a nutator mixer until the cell debris disappears. Insoluble polysaccharides, proteins, and lipids were pelleted at 10,000 x g for 5 minutes, and the supernatant was collected. DNA pellet was precipitated from the DNAzol by addition of ethanol in 2 parts DNAzol, 1 part ethanol ratio, and mixed very gently until no mixing lines could be seen. Spool DNA with a glass rod, however if DNA would not spool, either the lysozyme digestion step failed or the DNAzol added was insufficient for the size of the pellet; spin down the cell debris and add lysozyme solution again). The DNA pellet was washed three times with 10 ml of 75% ethanol and dried for 5 minutes. The DNA pellet was solubilized by addition of 500 µl of 8 mM NaOH and mixed thoroughly with a p1000 pipette, followed by 500 µl of TE (10 mM Tris HCl, 1 mM EDTA, pH 7.5). DNA concentration and quality in the supernatant was estimated by optical density measurements at OD_{260}/OD_{280}.

DNA Restriction Enzyme Digestion

DNA was digested by EcoR I (Invitrogen), EcoR I stock digests 15 units per µl, thus digests 15 µg DNA per hour. The volume of restriction enzyme added is calculated according to the size of the DNA pellet. The 10x H buffer was used for digestion at 37 ºC, and one-tenth volume of 10x H buffer was added to make the final volume at 1x concentration. Redo the optical density measurements to obtain a more accurate estimate of DNA concentration.

The steps of RNase treatment, cleanup, DNA solubilization and SYBR Green assay at this point were omitted from the current Bacterial Genomic DNA Hybridization Protocol.

Acrylamide Gel Electrophoresis

The 1D gel plates were assembled by siliconizing all the spacers, combs, and the notched glass plate, and the bottom and side spacers were clamped into place. 100 ml of 5% acrylamide solution (37.5:1 acrylamide: bis-acrylamide) was poured into a 250 ml filter flask. 100 µl of TEMED was added and sterile filtered, and after all solution was filtered it was swirled and allow for all bubbles to dissipate. It was followed by the addition of 200 µl of 20% APS, and the solution was swirled for mixing. The final loading dye was set at 1x concentration by mixing 10

µg of DNA with water and 5x loading dye. A stock solution of 50x 2DTAE was made by adding 2 M TrisBase (484.56 g), 1 M sodium acetate (164.06 g), 50 mM Na2EDTA (37.22 g), and 10% (v/v) Glacial acetic acid (200 ml), and bring the total volume to 2 L with double distilled water. The lower running tank was half-filled with 1x 2DTAE. The bottom spacer of the 1D gel plate was removed, and the lanes were marked with a marker on the plate with the smooth rounded edge. The 1D gel plate was inserted at an angle to remove bubbles. The aluminum plate was clamped onto the 1D gel plate, and the 1D gel plate was clamped onto the stand, the stand was raised to avoid letting the aluminum plate touch the buffer and screwed the stand in place. The top tank was filled with 1x 2DTAE and the samples were loaded into the wells. The gel was run at the setting of 0.1 amp, 265 V for 6 hours (or 100 V for 17 hours). The gel was stained in SYBR Green I which the 10,000x stock was diluted with 1x2DTAE down to 5x concentration. For the scanning step the phosphoimager aperture was set to 5.6, zoomed to 200, filtered to SYBR Green, and scan time was set to 20 minutes.

Random Primed DNA Labelling/Hybridization

Following final cleanup, digested DNA was labelled with Cy3 (GE Healthcare) using the Random Primed DNA Labelling Kit (Roche, Laval, PQ). 3 µg DNA (assayed by the SYBR Green I method) in 12.4 µl water was denatured at 95 ºC for 10 minutes and snap cooled on ice. 4.6 µl of Cy3 dye-dCTP/dNTP mix (0.4 µl of 1 mM Cy3 dye-dCTP, 0.4 µl water, 0.8 µl of 0.5 mM dCTP, 1 µl of 0.5 mM dATP, 1 µl of 0.5 mM dTTP, and 1 µl of 0.5 mM of dGTP) was added to the DNA, together with 2 µl hexanucleotide primer mix and 1 µl Klenow enzyme (2 Units). The DNA was labelled at 37 ºC for 1 hour, and stored at -20 ºC until hybridization.

Microarray slides were spotted with oligonucleotides from the Operon Genetics Human Genome Oligo Set Version 2.0 (Huntsville, AL). Slides were prepared under contract at the Prostate Cancer Research Centre Microarray Facility, at Vancouver General Hospital (Vancouver, BC), under the direction of Dr. Colleen Nelson. For hybridization to the microarray, 3.5 ml of hybridization buffer, made with 10% (w/v) Dextran sulphate (EMD Chemicals, Gibbstown, NJ), 1 M NaCl, 1% (v/v) TWEEN® 20, were added to the labelled DNA, and heated to 95 ºC for 3 minutes. Hybridization was conducted at 45 ºC overnight in the Fisher CS002551 – CMT hybridization chamber. After hybridization, slides were immediately rinsed with 1x SSC (0.15 M NaCl, 0.015 M sodium citrate) with 1% TWEEN 20 at room temperature for 10 minutes. Afterwards, the slides were washed in 0.5x SSC with 0.5% TWEEN 20 at room temperature for 2 minutes. The slides were then spin dried using a microarray chip spinner.

Microarray Digitization

After the microarray slides have dried, they were scanned at 550 nm excitation with 580 nm emission by the MACROview™ software in the DNAScope IV array scanner (Biomedical Photometrics, Waterloo, ON). Microarray image digitization was done using the GenePix Pro™ software from Axon Genetics (Palo Alto, CA), and the images were then saved as TIFF files for further evaluation. Within GenePix Pro, the user loads the GenePix Array List files designed for different types of microarray chips, and properties such as the number of features in the array, the number of subgrids, feature dimensions and shape, and upper or lower boundaries for maximum or background values could be adjusted as well. The GenePix Pro software automatically aligns the defined grid from the array list to the imaged microarray. The feature alignments were

checked one-by-one and manually refined by the user. These digitized values for each feature on the image were then extracted and exported as an Excel-readable flat text file for further data analysis. Samples from each species were organized by column and the different features by rows in the spreadsheet.

Image Analysis

All the hybridized microarray images were analyzed by the comparison of the hybridization patterns. The GenePix Pro software yields intensity values (average intensity over a fixed number of pixels) to three decimal places in a range from 0 to 65534. Once the data has been organized into spreadsheets in Microsoft Excel, all the zeros on the spreadsheet were replaced with ones to simplify the analysis and statistical comparisons, and the values were pruned to integer values in order to reduce data complexity.

For binning and data visualization, a tool called ChromaBlast in the BioTools add-in software in Excel was used. ChromaBlast determines a chip-dependent minimum and maximum value, then sorts that chip data into user-defined subsets (bins) of equal size. The number of bins (and the heat map colour assigned to the bins) is determined by the user. In this work, the range of values on the chip is from 0 to 65534, and 20 bins were set. Each bin is of equal size in terms of the total range, thus the bins in this work are: 1 to 3277, 3278 to 6553, 6554 to 9830, 9831 to 13107, 13108 to 16384, 16385 to 19660, 19661 to 22937, 22938 to 26214, 26215 to 29490, 29491 to 32767, 32768 to 36044, 36045 to 39320, 39321 to 42597, 42598 to 45874, 45875 to 49151, 49152 to 52427, 52428 to 55704, 55705 to 58981, 58982 to 62257, and 62258 to 65534. In this work, the same dark grey colour has been assigned to the lowest three bin ranges to minimize the effect of low intensity features as background noises. The highest three bin ranges were assigned the same red colour as well in order to differentially consider these higher-value features which were less prone to random variation. The differences in feature intensities were then visually compared on the "hot/cold" coloured heat map. Using the "data-filter" function in Excel, non-differentiating data such as large portions on the map in grey could be removed from display to simplify the analysis. After the data has gone through the binning process in ChromaBlast, the data were effectively normalized and would not required further numerical manipulations. In the future when large scale comparisons to an established library of species and strains are needed, then data normalization would be required.

Development of a Multipathogen Affymetrix DNA Microarray

In order to assess the utility of DNA microarrays for identification of bacterial pathogens to the species and strain level, a custom modified tiling multipathogen chip was designed for the Affymetrix platform. Organisms to include on the chip were derived from the National Institute of Allergy and Infectious Diseases list of priority pathogens. (http://www3.niaid.nih.gov/topics/BiodefenseRelated/Biodefense/research/CatA.htm). All Category A and B bacteria were selected in addition to *haemophilus influenzae*, *acenetobacter baumannii*, *chaetomium* species, *rickettsia* species, plasmids pBC16 and pLS1. Probes for viral pathogens were not included in this chip. Sequences selected for probe design obtained from regions that differed between strains of the same species especially those from Category A bacteria, regions that were constant within a species but differed between species, virulence genes and antimicrobial resistance genes.

To select sequences for probe selection, existing literature on bacterial microarray genotyping and strain differentiation and online databases were reviewed. The NCBI Protein Clusters database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=proteinclusters) was used to identify variants between strains. When variants were found, the DNA sequence was obtained by accessing the sequence viewer then strains that matched the sequence were identified by blasting (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against both the nucleotide references sequences then whole-genome shotgun sequences. For strains that differed by single base pair variations, SNP probes were constructed. These probes were 49 bases in length with the variant, designated by an "n", in the 25th (central) position. Antibiotic resistance genes were obtained from the Antibiotic Resistance Genes Database (http://ardb.cbcb.umd.edu/). The majority of the sequences used for probe selection were obtained from VFDB (http://www.mgc.ac.cn/VFs/main.htm), the Virulence Factors of Pathogenic Bacteria database. This database provided FASTA sequence of virulence genes and sequences that could be used for comparative genomics.

A master excel file was created in order to keep track of the 19,643 selected sequences. This file contained the probe name, organism used to obtain the sequence, gene ID/accession number/locus ID used to locate the gene, start and end base positions of the sequence used, length of the sequence segment used, first and last 8 bp of the selected sequence, gene name and description, strain the sequence was derived from and the complete sequence segment. From this master file, two files were prepared for Affymetrix. The first was the instruction file listing the probe name, start and end positions of the sequence, first and last 8 bp of the sequence and a description of the probe. The second file included all the sequences in FASTA format, each identified by the probe name provided in the instruction file. Once Affymetrix received these files, a maximum of five 25-mer probes were designed for each sequence. Degenerate probes were removed and a list of the proposed microarray design was returned for evaluation. This microarray was constructed using 57,061 probes from 11,516 unique microbial sequences, 24,660 probes from 264 SNP sequence and approximately 140,000 random probes along with controls to fill in 220,678-probe chip.

Results

A preliminary library of 19 microbial genomic DNA samples has been set up and archived for future use. When genomic DNA from more microbial strains have been prepared, along with the existing data which have not been analyzed yet, this library can be expanded further and serve as a reference for future work.

ChromaBlast, an add-in tool in Microsoft Excel, was used to normalize all the array data sets, and produced color heat maps which enable us to differentiate microbial genomic DNA samples based on the intensities. There are three different array data sets from this work, the first set obtained from microbial gDNA samples hybridized on Human 21K chips (Figure 1), the second set obtained from samples hybridized on Bacterial chips (Figure 2), and the third set obtained from the 10 samples with duplicates available hybridized on Human 21K chips (Figure 3). There are 23,232 and 5,280 features on the Human 21K chips and Bacterial chips, respectively. The first 1,000 features on the color heat maps are included in Figure 1-3. The columns representing microbial gDNA samples are arranged on the color heat maps differently for each data set, usually by grouping the related bacterial species and strains next to each other.

In Figure 4-5, two sets of features which shown significant differences in intensities between the strains are demonstrated. In Figure 4 with the Human 21K array data set, at feature number 5333 the intensities are higher for *G. stearothermophilus* and the *Bacillus* species comparing to the *E. coli TOP10* strains, as evident by most of the *Bacillus* species being assigned the higher value bins in red and orange colors. On the other hand, at feature number 9354 the *E. coli TOP10* strains have higher intensities when compared to the *Bacillus* species, since most of the *Bacillus* species were assigned the lower value bins in blue colors. Similarly, in Figure 5 with the Bacterial array data set, at feature number 2512 the *E. coli TOP10* strains had higher intensities and were assigned the higher value bins when compared to the *Bacillus* species, while the opposite case was true at feature number 3686.

In both of the color heat maps generated from the Human 21K and Bacterial array data sets, features with the highest intensities across all the samples, plus those features with significant difference in intensities and deemed to be important in microbial species and strains differentiation are recorded in Table 1 and 2.

Different intensity patterns are found in the color heat maps between the samples, and they are useful in discriminating the different species and strains from each other. In this work, there are feature sites found on the color heat maps with enough information which allow us to differentiate the genera, from *Bacillus* species to the *E. coli TOP10* strains with greater confidence. No color heat map from the samples has been found to be exactly the same for any species or strains, therefore the potential to discriminate microbial species and strains do exist. However, with the limited data sets available the species and strain differentiation was determined with lower confidence at this stage.

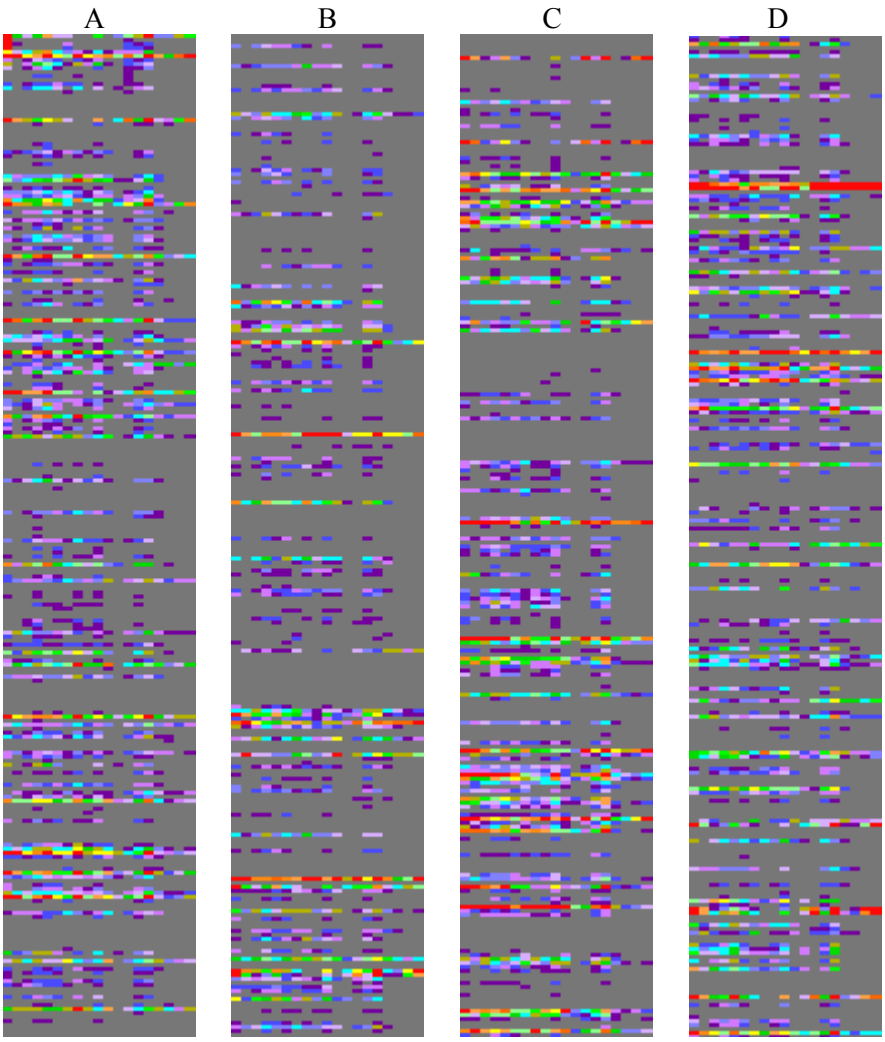

Figure 1: Analysis by ChromaBlast of first 1000 features of the Human 21K array data set from 19 different gDNA samples. Data are unfiltered.

Bin	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Color																				
Max Intensity	3277	6553	9830	13107	16384	19660	22937	26214	29490	32767	36044	39320	42597	45874	49151	52427	55704	58981	62257	65534
Count	65472	58800	37220	26519	19320	14599	11384	8835	6931	5560	4552	3794	2972	2356	1914	1525	1187	1008	843	1261

Column	Features
A	1-250
B	251-500
C	501-750
D	751-1000

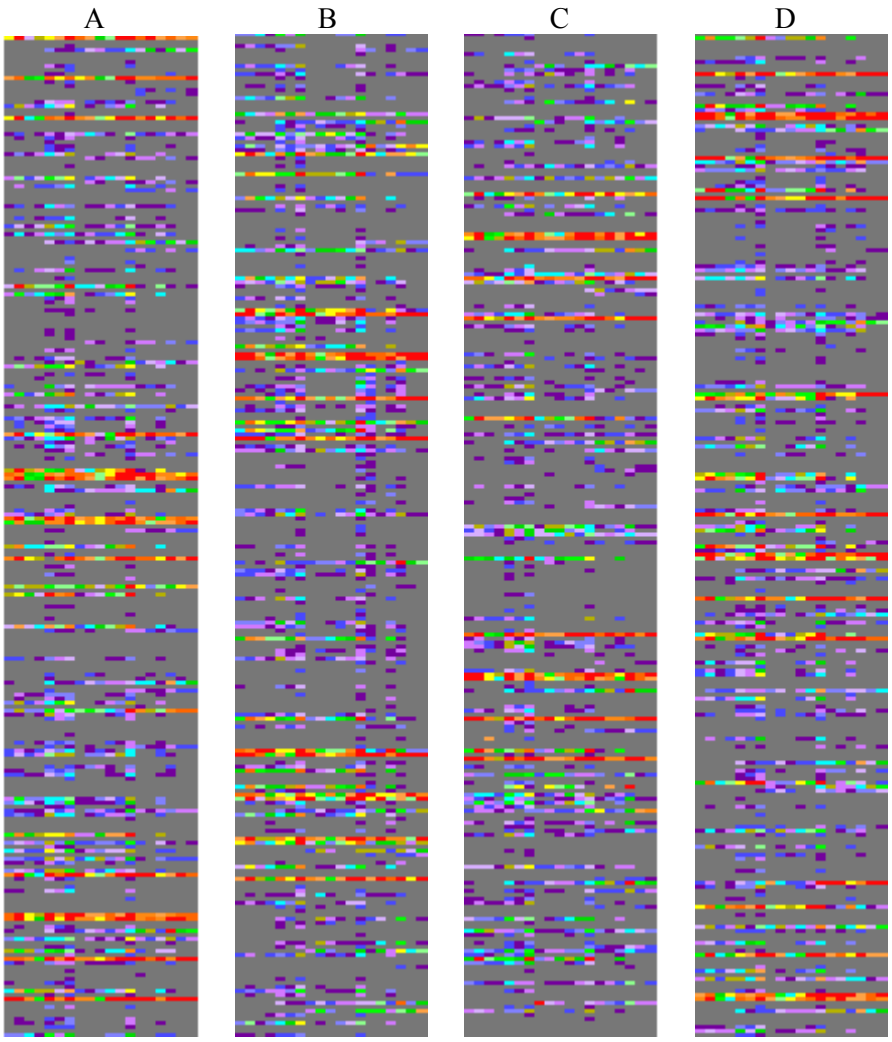

Figure 2: Analysis by ChromaBlast of first 1000 features of the Bacterial array data set from 19 different gDNA samples. Data are unfiltered.

Bin	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Color																				
Max Intensity	3277	6553	9830	13107	16384	19660	22937	26214	29490	32767	36044	39320	42597	45874	49151	52427	55704	58981	62257	65534
Count	47518	15146	8669	5775	4239	3108	2311	1883	1547	1248	1054	943	861	925	973	919	939	913	731	618

Column	Features
A	1-250
B	251-500
C	501-750
D	751-1000

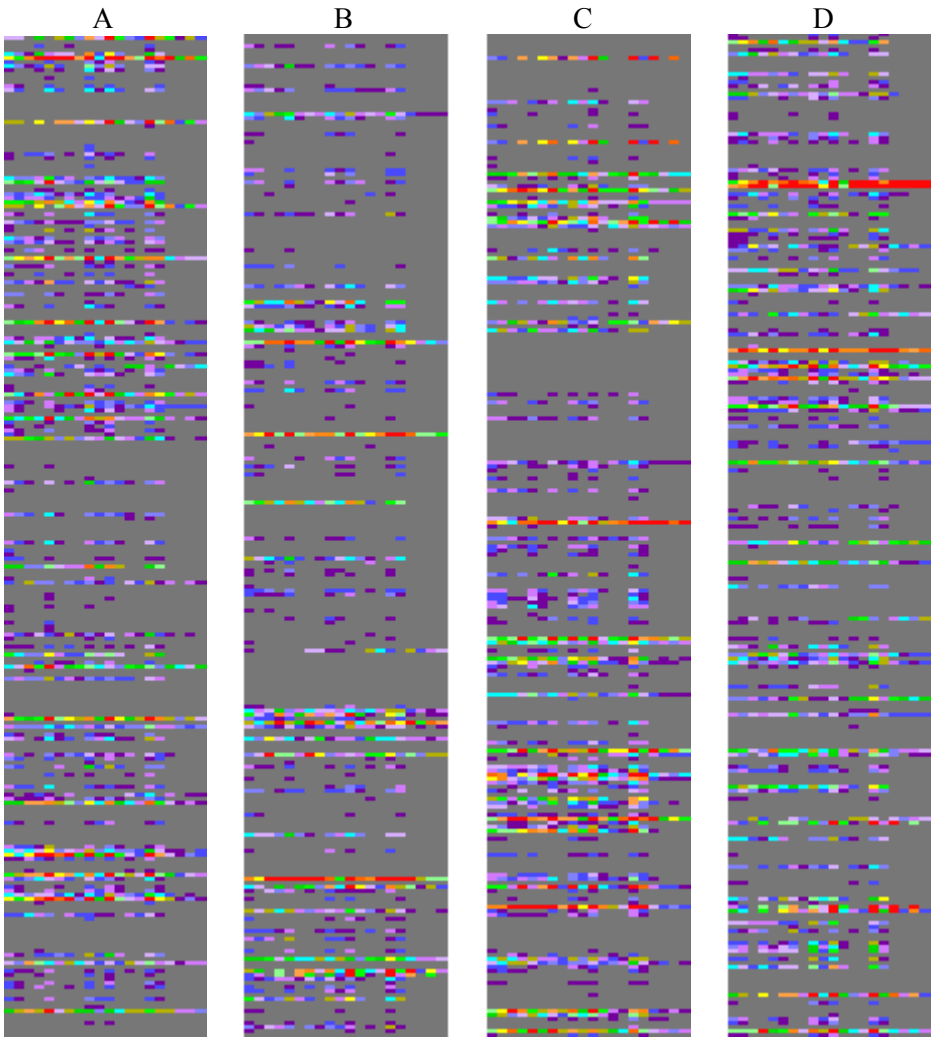

Figure 3: Analysis by ChromaBlast of first 1000 features of the Human 21K array data set from 10 strains with one duplicate each. Data are unfiltered.

Bin	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Color																				
Max Intensity	3277	6553	9830	13107	16384	19660	22937	26214	29490	32767	36044	39320	42597	45874	49151	52427	55704	58981	62257	65534
Count	65472	56378	35783	25297	18077	13601	10396	8067	6235	5118	4280	3420	2605	2111	1727	1406	1099	998	877	1432

Column	Features
A	1-250
B	251-500
C	501-750
D	751-1000

Feature #5333

Feature #9354

Figure 4: Comparison of intensities at two specific features on the Human 21K array. The data represented are (left to right): G. stearothermophilus, B. subtilis 168, B. megaterium, B. thuringiensis, B. thuringiensis kurstaki, B. circulans, B. amyloliquefaciens, B. sphaericus, B. mycoides, B. globigii, B. licheniformis, B. coagulans, E. coli TOP10 HK43D LTC(LTC/HD)-60-1(3), E. coli TOP10 HK43D LTC(-LTC/HD)-60-1(5), E. coli TOP10 HK43D LTC/HD(-LTC)-40-1(9), E. coli TOP10 HK43D LTC/HD(-LTC)-40-1(6), E. coli TOP10 HK43D LTC/HD(-LTC)-40-1(11), E. coli TOP10 HK43D LTC/HD-C-40-1(3), E. coli TOP10 HK43D LTC/HD-C-40-1(1).

Feature #2512

Feature #3686

Figure 5: Comparison of intensities at two specific features on the Bacterial array. The data represented are (left to right): B. sphaericus, B. mycoides, B. globigii, B. coagulans, B. circulans, B. thuringiensis, B. thuringiensis kurstaki, B. subtilis 168, B. amyloliquefaciens, B. licheniformis, B. megaterium, G. stearothermophilus, E. coli TOP10 HK43D LTC/HD(-LTC)-40-1(9), E. coli TOP10 HK43D LTC/HD(-LTC)-40-1(11), E. coli TOP10 HK43D LTC/HD(-LTC)-40-1(6), E. coli TOP10 HK43D LTC(-LTC/HD)-60-1(5), E. coli TOP10 HK43D LTC(LTC/HD)-60-1(3), E. coli TOP10 HK43D LTC/HD-C-40-1(3), E. coli TOP10 HK43D LTC/HD-C-40-1(1).

Table 1: Notable Features on the Color Heat Map for the Human 21K array.

Features with Differentiable Intensities across the Samples	Features with High Intensities across all Samples
491, 656, 814, 877, 916, 1040, 1081, 1151, 1221, 1368, 1443, 1573, 1589, 1617, 1715, 2130, 2155, 2535, 2708, 2743, 2805, 2912, 3093, 3558, 4368, 4698, 4778, 4946, 5027, 5164, 5192, 5226, 5333, 5398, 5527, 5802, 5856, 5892, 5951, 6001, 6447, 6460, 6479, 6560, 6583, 6611, 6693, 6715, 6806, 6808, 6819, 6828, 6850, 6875, 6879, 6889, 6890, 6939, 7254, 7272, 7340, 7418, 7517, 7590, 7593, 7671, 8478, 8533, 8717, 8982, 9321, 9354, 9459, 10056, 10175, 10544, 10723, 10748, 10843, 10879, 10897, 10908, 11090, 11464, 11703, 11741, 11803, 11805, 11807, 12150, 12268, 12378, 12382, 12446, 12458, 12524, 12708, 12788, 14682, 14752, 14834, 14942, 15271, 15869, 16021, 16066, 16102, 16187, 16265, 16514, 16533, 16632, 16660, 16909, 17139, 17187, 17433, 17439, 17480, 17534, 17619, 18122, 18550, 18555, 18604, 18773, 19167, 20566, 22534, 22624, 22718, 22739, 22866	622, 787, 788, 1114, 1304, 2351, 2490, 2611, 2627, 2784, 2979, 3240, 3322, 3398, 3679, 3702, 3741, 3948, 4181, 4494, 4603, 4647, 5400, 6340, 6548, 6665, 6849, 6925, 6988, 7152, 7258, 7680, 7937, 8292, 8613, 8790, 8890, 9516, 9788, 10209, 10374, 10550, 10852, 11075, 11189, 11368, 11577, 11812, 11954, 12176, 12225, 12444, 12730, 12865, 12868, 13008, 13051, 13666, 14358, 14606, 15074, 15100, 15191, 15617, 16278, 16387, 17233, 17677, 17749, 17976, 18026, 18028, 18238, 18861, 20056, 21995, 22429, 22722

Table 2: Notable Features on the Color Heat Map for the Bacterial Array.

Features with Differentiable Intensities across the Samples	Features with High Intensities across all Samples
63, 109, 128, 140, 169, 200, 224, 278, 285, 319, 328, 382, 421, 429, 494, 554, 596, 602, 664, 694, 712, 742, 751, 768, 774, 789, 822, 860, 900, 962, 968, 1055, 1083, 1112, 1200, 1214, 1269, 1402, 1433, 1443, 1482, 1599, 1654, 1665, 1702, 1813, 1922, 1931, 2032, 2036, 2062, 2102, 2120, 2142, 2151, 2194, 2212, 2218, 2307, 2391, 2446, 2512, 2560, 2590, 2634, 2639, 2696, 2802, 2806, 2825, 2840, 2886, 2974, 2978, 3214, 3223, 3295, 3303, 3325, 3365, 3463, 3500, 3553, 3798, 3822, 4003, 4128, 4186, 4294, 4303, 4351, 4370, 4760, 5051	1, 11, 21, 110, 111, 121, 122, 210, 220, 221, 231, 241, 320, 330, 331, 334, 341, 351, 430, 550, 551, 561, 571, 650, 660, 661, 671, 681, 760, 770, 771, 781, 791, 840, 870, 891, 901, 980, 990, 991, 1001, 1011, 1090, 1100, 1101, 1111, 1121, 1210, 1211, 1221, 1231, 1250, 1302, 1310, 1320, 1331, 1341, 1360, 1394, 1420, 1430, 1431, 1441, 1451, 1489, 1530, 1540, 1541, 1551, 1552, 1561, 1572, 1640, 1650, 1651, 1664, 1718, 1750, 1771, 1781, 1800, 1870, 1871, 1881, 1891, 1944, 1962, 1970, 1980, 1981, 1991, 2001, 2065, 2090, 2091, 2096, 2101, 2104, 2110, 2111, 2118, 2129, 2144, 2158, 2184, 2190, 2200, 2201, 2211, 2221, 2228, 2300, 2310, 2311, 2321, 2331, 2392, 2410, 2420, 2421, 2426, 2431, 2441, 2490, 2494, 2501, 2520, 2530, 2531, 2541, 2551, 2630, 2640, 2641, 2651, 2661, 2750, 2751, 2761, 2771, 2803, 2823, 2842, 2850, 2860, 2861, 2870, 2871, 2874, 2881, 2898, 2899, 2914, 2919, 2970, 2971, 2981, 2991, 3031, 3062, 3066, 3070, 3080, 3091, 3101, 3196, 3201, 3211, 3263, 3273, 3290, 3311, 3321, 3385, 3392, 3400, 3410, 3411, 3431, 3495, 3510, 3520, 3595, 3620, 3641, 3686, 3751, 3761, 3840, 3850, 3871, 3919, 4008, 4145, 4311, 4631, 4641, 4721, 4914, 5070, 5081, 5183, 5191, 5210, 5280

Discussion

With thousands of DNA probe sequences spotted onto microscope slides, many features were assayed in a single sample simultaneously. One of the ways microarray technology could be utilized to complement existing identification technologies is to use DNA probes which are specific to known microbial sequences. As the result of the specific hybridizations using these probes, the species and strain of the microbes could be identified with high confidence. However, if the microbe in question is a recombinant containing novel genetic content, identification will be problematic since the requirement of known sequences on these recombinants cannot be met. Another approach would be to use non-specific oligonucleotides as probes, which will not bind to any known microbial sequences on the panel and produce hybridization patterns that are unique for each microbe. The patterns for the same samples used are reproducible, therefore genomic DNA are hybridized to the features is sequence specific ways.

The step of manually refining feature alignments by the user in microarray digitization is found to be the most time consuming process in microarray. There are numerous microarray chips produced for each microbial species and strains, unfortunately no more than two from each strain have been included in the data sets due to time restriction. Besides adding more microbial species and strains to the library in future microarray work, existing data available for further analysis should be included as well.

Some of the advantages in utilizing microarray are that sample DNA amplification is not necessary for most biological threat agents. In the study of recombinant microbes, instead transporting the unknown samples with biohazard potential, microarray slides themselves are safe for transportation and the protocols and equipment needed for the experiments are commercially available. Given the library is robust, in a single experiment a test sample could be compared to hundreds of known strains and be identified promptly, or its closest match would be found and the degree of similarity to known strains could be estimated. Furthermore, if DNA probes that are specific for genes encoding additional characteristics such as antibiotic resistance, toxin or virulence are included, microbes with these genetically modified traits which are potentially more harmful would be identified as well. With this knowledge in mind, the appropriate decisions for personal protection, medical responses, or quarantine could be made.

References

[1] Song, L., Ahn, S., Walt, D. (2005). Detecting Biological Warfare Agents. Emerging Infectious Diseases, 11(10): 1629-1632.

[2] Charbonnier, Y., Gettler, B., Francois, P., Bento, M., Renzoni, A., Vaudaux, P., Schlegel, W., Schrenzel, J. (2005). A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus*. BMC Genomics, 6: 95.

[3] Pan, W., Lin, J., Le, C. (2003). A mixture model approach to detecting differentially expressed genes with microarray data. Functional & Integrative Genomics, 3: 117-124.

[4] Cummings, C., Relman, D. (2000). Using DNA microarrays to study host-microbe interactions. Emerging Infectious Diseases, 6(5): 513-525.

[5] Rubins, K., Hensley, L., Jahrling, P., Whitney, A., Geisbert, T., Huggins, J., Owen, A., LeDuc, J., Brown, P., Relman, D. (2004). The host response to smallpox: Analysis of the gene expression program in peripheral blood cells in a nonhuman primate model. PNAS, 101(42): 15190-15195.

[6] Eisen, M., Spellman, P., Brown, P., Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. The Proceedings of the National Academy of Sciences USA, 95: 14863-14868.

[7] Burton, J., Oshota, O., North, E., Hudson, M., Polyanskaya, N., Brehm, J., Lloyd, G., Silman, N. (2005). Development of a multipathogen oligonucleotide microarray for detection of *Bacillus anthracis*. Molecular and Cellular Probes, 19: 349-357.

[8] Barrett, M., Scheffer, A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P., Yakhini, Z., Bruhn, L., Laderman, S. (2004). Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. PNAS, 101(51): 17765-17770.

[9] Wang, D., Coscoy, L., Zylberberg, M., Avila, P., Boushey, H., Ganem, D., DeRisi, J. (2002). Microarray-based detection and genotyping of viral pathogens. PNAS, 99(24): 15687-15692.

[10] Goldberg, C., Wang, E., Yi, C., Goldberg, T., Brando, L., Marincola, F., Webster, M., Torrey, E. (2005). Infectious pathogen detection arrays: viral detection in cell lines and post-mortem brain tissue. BioTechniques, 39: 741-751.

[11] Dudoit, S., Gentleman, R., Quackenbush, J. (2003). Open source software for the analysis of microarray data. BioTechniques, 34: S45-S51.

[12] Lefkowitz, E. (2003). Development of a viral biological-threat bioinformatics resource. The University of Alabama at Birmingham, Annual Report.

[13] Workman, C., Jensen, L., Jarmer, H., Berka, R., Gautier, L., Nielsen, H., Saxild, H., Nielsen, C., Brunak, S., Knudsen, S. (2002). A new non-linear normalization method for reducing variability in DNA microarray system. Genome Biology, 3(9): 1-16.

List of symbols/abbreviations/acronyms/initialisms

APS	Ammonium persulfate
dNTP	dATP, dCTP, dGTP, dTTP : deoxynucleotide triphosphates of DNA bases
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetraacetic acid
OD_{260}/OD_{280}	Ratio of optical densities at 260 and 280 nm
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
SSC	Salt, sodium citrate
TEMED	Tetramethylethylenediamine
TIFF	Tagged Image File Format
TRIS	2-amino-2-hydroxymethyl-1,3-propanediol
TWEEN 20	Polyoxyethylene (20) sorbitan monolaurate

Glossary

Antibiotic Resistance

The ability of a microbe to withstand the effects of antibiotics, allows it to survive and grow in the environment which would normally inhibit its growth. It could occur naturally via natural selection or introduced by artificial means. The resistance genes could be transferred to another microbe to give rise to novel organisms with such properties.

Fingerprint

A collection of signal intensity data sets from the digitized images of genomic DNA hybridized to a microarray spotted with DNA fragments. Such fingerprint of any given microbial species or strain is unique to the specific microbe.

Gene

A locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and other functional sequence regions.

Genomic DNA

The DNA which makes up the genome of an organism and contains the genetic information for all proteins produced by an organism.

Hybridization

Fluorescent labelled sample DNA is applied to the microarray surface, after an incubation period under controlled temperature and solution concentrations, sequences in the sample DNA complementary to the sequences in the microarray will bind together. Hybridization refers to the entire process from labelling to binding to the final washing steps.

Library

A collection of DNA which is maintained as a source of genetic material. Or alternatively, a data set of digitized images from microarray assays used for comparison purpose.

Microarray

An arrayed series of thousands of microscopic spots of DNA oligonucleotides (features), each cotaining picomoles of a specific DNA sequence. The probes are attached to a solid surface by a covalent bond to a chemical matrix. The DNA could come from cDNA libraries, PCR products, genomic DNA, or synthetic oligonucleotides.

Nucleotide

The structural units of RNA and DNA, which encodes genetic information by the order of the nucleotides in the chain. Organic compounds which consists of three joined structures: a nitrogenous base (purines or pyrimidines), a sugar (ribose for RNA, deoxyribose for DNA), and a phosphate group.

Oligonucleotide

A short segment of RNA or DNA, and they are often used as probes for detecting DNA or RNA because they bind readily to their complements. Oligos may be random in sequence so the sequence does not derive from known DNA sequences.

Recombinant

An organism containing novel or additional genetic materials, either by natural or artificial means.

Restriction Enzyme

An enzyme that cuts double-stranded DNA following its specific recognition of short nucleotide sequences, known as restriction sites, in the DNA.

Species

Microbes that are grouped together according to major genetic differences.

Strain

A genetic variant or subtype of a microbe, with slight differences from the other members of the same species (e.g. antibiotic resistance).

Toxin

A gene product which is poisonous to the host organism or other microbes once it is expressed.

Virulence

The degree of pathogenicity of a microbe, or the relative ability of a microbe to cause disease.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Canada West Biosciences Inc.
5429 60th Street
Camrose, AB T4V 4G9

2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Microarray Genomic Systems Development

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Lam, V.; Crichton, M.; Dickinson Laing, T.; Mah, D.C.

5. DATE OF PUBLICATION (Month and year of publication of document.)

June 2008

6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)

30

6b. NO. OF REFS (Total cited in document.)

13

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

DRDC Suffield, PO Box 4000, Station Main, Medicine Hat, AB, T1A 8K6

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

W7702-04-R053/001/EDM

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)

DRDC Suffield CR 2009-145

10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

In order to identify and characterize microbes using the currently available methods, information on the organism's genetic material is required. Recombinant microbes which contain novel genetic material would not have such information and therefore avoid detection. Previous work has been done to develop a system with array-based genomic fingerprinting technology, which should help to identify the new strains, detect the presence of novel genetic material, and measure gene expressions. The next phase is the development of a test system for rapid species identification in addition to the oligonucleotide fingerprint. The test organisms for this work were *Bacillus* bacteria (11 species), *Escherichia coli TOP10* (7 strains), and *Geobacillus stearothermophilus.* Using standard molecular biology methods, we isolated genomic DNA, digested the DNA to reduce complexity, labelled it with fluorescent dyes, and hybridized the labelled DNA to two types of microarrays, Human Operon 21K chips containing 23,232 features and Bacterial genomic chips containing 5,280 features. The hybridization data was then analyzed with ChromaBlast, an useful analytic tool in Excel, which normalized columnar data, sorted the data into user-selectable range-driven bins, developed colour heat maps from the data, and then output the heat map and bin assortment for review. When the data patterns on the colour heat maps were filtered and sorted, bacteria in different genera could be discriminated with high confidence in certain subsets of the features. However, species or strain differentiations of the test organisms were not as evident in this work. A multipathogen chip was designed to further investigate species and strain differentiation. Due to time limitation on the contract, only one or two sample chips from each test organism has been included thus far. In the next phase, data from more replicate microarray chips should be included in the set of hybridization data.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

ChromaBlast; Genomic Fingerprinting; Hybridization; Microarray

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

www.drdc-rddc.gc.ca